Schizophrenia Bulletin vol. 37 no. 3 pp. 456–463, 2011
doi:10.1093/schbul/sbr031

Schizophrenia Genetics: Where Next?

Yunjung Kim[1], Stephanie Zerwas[2], Sara E. Trace[2], and Patrick F. Sullivan[*,1,2]

[1]Department of Genetics, University of North Carolina, Chapel Hill, NC; [2]Department of Psychiatry, University of North Carolina, Chapel Hill, NC

*To whom correspondence should be addressed; Department of Genetics, CB#7264, 5097 Genomic Medicine, University of North Carolina, Chapel Hill, NC, 27599-7264, US; tel: +919-966-3358, fax: +919-966-3630, e-mail: pfsulliv@med.unc.edu.

The purpose of this invited review is to summarize the state of genetic research into the etiology of schizophrenia (SCZ) and to consider options for progress. The fundamental uncertainty in SCZ genetics has always been the nature of the beast, the underlying genetic architecture. If this were known, studies using the appropriate technologies and sample sizes could be designed with an excellent chance of producing high-confidence results. Until recently, few pertinent data were available, and the field necessarily relied on speculation. However, for the first time in the complex and frustrating history of inquiry into the genetics of SCZ, we now have empirical data about the genetic basis of SCZ that implicate specific loci and that can be used to plan the next steps forward.

Key words: schizophrenia/genetics/review/genome-wide association/genome-wide linkage/next-generation sequencing

What are the Goals of Genetic Studies of SCZ?

The major goal of SCZ genetic research is to develop a complete list of genetic loci and pathways that confer risk or protection. Given how little we know for certain about this enigmatic, clinically heterogeneous, and genetically complex disorder, even one unequivocal insight would be of immense value. The enormous advantage of genetic studies compared with virtually all other human biomarker studies is that the key element of causation is present: exposure to a genetic risk factor begins at conception and prior to disease onset, and thus temporality is satisfied.

The second but more distal goal requires completion of the first: to reduce SCZ incidence, to minimize time-to-treatment for incident cases, and to reduce morbidity and mortality in cases with SCZ. We stress that this is an ultimate and not a proximal goal.

Where Have We Been?

A number of reviews of the history of genetic approaches to SCZ are available.[1] Table 1 summarizes the main approaches. Prior to 2008, 6 methodological approaches were predominant.

First, generations of physicians have evaluated the family histories of probands with SCZ. It is notable that this informal surveillance network has not identified any pedigree with SCZ segregating in a Mendelian fashion. This stands in contrast to many other complex traits where Mendelian subforms have been identified. For example, small proportions of cases with breast cancer, Alzheimer's disease, and type 2 diabetes mellitus (T2DM) are caused by single gene mutations that have very high penetrance and are typified by early age of onset. Notably, genetic evaluation of childhood-onset SCZ cases has not yielded causal mutations.[17] Because the clinical surveillance network has been nonsystematic, the main conclusion can only be that SCZ is unlikely to have Mendelian subforms.

Second, a large body of genetic epidemiological studies has provided the rough outlines of the genetic architecture of SCZ. There is strong but indirect evidence that SCZ is familial and highly heritable. Third, application of segregation analysis to SCZ pedigree data has been inconclusive. Consistent with inference from the clinical surveillance network, it is possible to reject a few extreme models (eg, exclusively dominant or recessive Mendelian models). However, many other genetic models are consistent with the data.

Fourth, evaluation of "microscopic" genomic changes using cytogenetic methods has been informative. The most notable finding was the identification of the 22q11.2 deletion, a rare, potent, and nonspecific risk factor for SCZ. Its prevalence in cases is ~0.3% with

Table 1. Genetic Approaches to SCZ and Broad Conclusions

Method	Basis	Conclusions
Clinical surveillance	The clinical process of physicians taking family histories. No DNA analysis.	Over past 100 years, no pedigrees with unequivocal Mendelian inheritance reported
Genetic epidemiology	Diagnosis of SCZ in various types of relatives (family, twin, adoption). No DNA analysis.	Family history an important SCZ risk factor Most SCZ sporadic (>90%) SCZ familial: λ_{sibs} 8.6 (7.9–9.6) SCZ heritable: 81% (73–90%)
Segregation analysis	What inheritance models are statistically consistent with observed diagnoses of SCZ in pedigrees? No DNA analysis.	Uninformative, many different models consistent with data
Cytogenetics	Search for genomic changes in SCZ cases using karyotyping (3 Mb or larger)	Deletion of 22q11.2 is a rare, potent, and nonspecific risk factor for SCZ
Genome-wide linkage	Attempt at an unbiased genome search for regions whose inheritance in pedigrees is correlated with SCZ	No findings meeting modern criteria for significance and replication
Candidate gene association	Evaluation of frequencies of genetic variation in SCZ cases vs controls for genes selected using prior hypotheses about SCZ etiology	Major research focus but no findings meeting modern criteria for significance and replication, most studies with serious methodological issues
Genome-wide association	Attempt at unbiased genome search for loci with differing frequencies in SCZ cases vs controls	Multiple regions meeting modern criteria for significance and replication Evidence that SCZ is highly polygenic No common variants of strong effect
CNVs	Search for genomic changes in SCZ cases vs controls (resolution variable but can be 100 kb or smaller)	Multiple regions that are rare, relatively potent, and nonspecific risk factors for SCZ (22q11.2, 15q13.3, 16p11.2, 1q21.1, *NRXN1*)
Resequencing	Use high-throughput methods to resequence regions, exomes, or genomes to identify SNPs, indels, and CNVs with differing frequencies (individually or in aggregate) in SCZ cases vs controls	Initial studies in progress

Note: Literature cited: genetic epidemiology,[2–5] segregation analysis,[6,7] cytogenetics,[8] genome-wide linkage,[9] candidate gene association (Collins et al, Submitted,[10,11] genome-wide association (Schizophrenia Psychiatric Genome-Wide Association Study Consortium, Submitted[12]), and CNVs.[13–16]

genotypic relative risk of ~20.[18] This deletion does not act in a Mendelian fashion: a minority of individuals with this deletion develops SCZ and it increases risk for multiple other neuropsychiatric disorders along with the somatic features of velo-cardiofacial syndrome.[19]

Fifth, there have been over 30 genome-wide linkage studies of SCZ along with a meta-analysis.[9] No genomic region emerged that exceeded genome-wide significance and which was reproducible across studies. These results are typical for complex biomedical diseases.

Sixth, hypothesis-driven candidate gene studies have been a major focus in SCZ research with the SZGene database listing >1400 studies since 1965[20] (for comparison, there are ~2200 PubMed citations for "SCZ randomized controlled trials"). This body of work has not yielded associations that meet modern criteria for replication.[21] Indeed, the nonsystematic nature of this search has led to mistakes (eg, *TCF4* has strong evidence of association with SCZ and yet has multiple negative studies of the wrong variant). There are major problems with the hypothesis-driven candidate gene approach.

- The information content of our hypotheses about SCZ is unclear. It is possible that many ideas that have dominated the field are at least partially incorrect. More generally, high-confidence genetic results have poor correspondence with prior hypotheses about etiology for most complex traits. T2DM has a wealth of advantages over SCZ (a biochemically defined case definition, Mendelian subforms, accessible human tissues, valid animal models, more sophisticated knowledge of its biology, greater funding, and more research groups) and yet the ideas that guided hypothesis-driven candidate gene studies correspond poorly with the current list of genes very strongly implicated in its etiology. SCZ has every disadvantage compared with T2DM: is it plausible that our hypotheses about SCZ are more likely to correspond to the genetic underpinnings?

- Multiple comparison control (type 1 error) requires scrupulous attention to enumerating every hypothesis tested.[22] This is easy to do in genome-wide association studies (GWAS) but may not have been a general feature in the candidate gene literature.
- Statistical power (type 2 error) is generally very low in the SCZ hypothesis-driven candidate gene literature. Even for relatively large studies (ie, samples sizes at the 90th percentiles of N_{case} = 537 and $N_{control}$ = 628 from the SZGene database), a liberal correction for multiple comparisons (α = 0.005, 10 markers), and an implausibly large effect size (median genotypic relative risk = 1.28 and median minor allele frequency of 0.29),[23] power was only 48%.
- The SZGene database contains records for 732 autosomal genes from 1374 hypothesis-driven candidate gene studies. These genes were studied from 1–81 times but most genes (563, 76.9%) were investigated in 1 (60.9%) or 2 studies (16.0%). Although replication is critical in human genetics,[21] there is little capacity to evaluate both false positive and false negative findings.
- Genetic variation was typically poorly captured. For technological or cost reasons, one or a few genetic markers were assessed per gene. This approach cannot capture even the common variation known to be present nearly everywhere in the genome.

In sum, prior to 2008, the cytogenetic finding of an association of SCZ with the 22q11.2 deletion was the only robust and reproducible genetic association for SCZ.

What Do We Know Now?

GWAS have yielded a plethora of findings[23,24] that meet modern criteria for replication in human genetics.[21] A "primer" is available.[25] Since 2005, >700 GWAS have been published; considering findings exceeding a conservative significance threshold ($P < 5 \times 10^{-8}$), GWAS have implicated ~1500 genetic markers for 101 human diseases and 124 biomedical traits (eg, height, body mass index (BMI), and lipid levels). GWAS have produced more etiological knowledge than virtually any other technology in the history of medicine, save for clinical microbiology and radiology.

There are 8 published SCZ GWAS of European samples that used individual-level genotyping[26–33] (table 2). In addition, 2 studies included African-American samples,[31,33] one subjects of Japanese ancestry,[34] and 2 used less reliable DNA pooling methods.[35,36] By current standards in human genetics, the sample sizes for virtually all of these studies are small. Therefore, the Psychiatric GWAS Consortium[37] has conducted an integrated mega-analysis of all available GWAS data on European samples (Schizophrenia Psychiatric Genome-Wide Association Study Consortium, Submitted[12]).

In contrast to the paucity of findings from prior methods, GWAS has "worked" for SCZ in terms of identifying common genetic variation that meet modern standards for replication and significance in human genetics. These findings include:

- Genetic variation in the extended major histocompatibility complex (MHC) locus on chromosome 6 is associated with SCZ.[27,31,32]
- Genetic variation near *MIR137* (the gene encoding the microRNA miR-137) exceeds genome-wide significance. At least 4 other genes exceeding genome-wide significance contain predicted miR-137 binding sites (Dr Stephan Ripke, oral presentation, World Congress on Psychiatric Genetics, Athens, Greece, 6 October 2010). Given the likely role of miR-137 in neuronal development and cross talk with epigenetic modification machinery, these GWAS findings point at a novel, plausible, and falsifiable hypothesis about the etiology of SCZ.
- Additional loci with strong support include *TCF4*, a region of chr10, *CSMD1*, *NGRN*, and, in joint analyses with bipolar disorder, *ZNF804A*, *CACNA1C*, and *ANK3*.
- The above loci reach a conservative threshold for genome-wide significance intended to minimize false positive claims. There is evidence, however, that there are hundreds and perhaps thousands of additional common loci that confer susceptibility to SCZ.[27] A polygenic model from a SCZ discovery set replicated in 3 independent SCZ samples and, critically, did not replicate in 6 nonpsychiatric biomedical disorders ruling out many types of bias. This polygenic model has replicated in a fourth independent sample of parent-affected offspring trios, effectively ruling cryptic population stratification.
- Copy number variants (CNVs) are rare but potent risk factors for SCZ. CNVs that meet stringent criteria are deletions in 22q11.21, 15q13.3, *NRXN1*, and 1q21.1 along with duplications in 16p11.2.[18] All of these are rare (present in 1–3 SCZ cases per 1000) but strong risk factors for SCZ (genotypic relative risks of 7–20). These large genomic changes are nonspecific, and increase risk for multiple psychiatric, neurological, and general medical disorders.[15]
- Several genetic architectures for SCZ can be excluded with high confidence. SCZ is not exclusively caused by rare variants of strong effect (ie, it is not a collection of many different Mendelian disorders), and common variants of strong effect in European samples are very unlikely to exist.[7,38]

The MHC finding has been criticized as being due to bias or artifact. However, the empirical results meet accepted criteria for replication in human genetics, the *P* value exceeds chance by 10 000×, there are consistent effects across samples, and appropriate control for population stratification does not explain away the

Table 2. GWAS for SCZ in European Samples Using Individual Genotyping

		Stage 1			Stage 2		
First author	Year	N_{cases}	$N_{controls}$	Significant regions	N_{cases}	$N_{controls}$	Significant regions
Lencz	2007	178	144	0	—	—	—
O'Donovan	2008	642	2937	0	7308	12 834	0
Sullivan	2008	417	411	0	—	—	—
ISC	2009	3322	3587	1	8008	19 077	2
Need	2009	900	877	0	1592	2114	0
Shi	2009	2681	2653	0	8008	19 077	2
Stefansson	2009	2663	13 498	0	12 945	34 591	4
Athanasiu	2010	201	307	0	2864	14 087	0
PGC	—	9394	12 462	2	17 836	33 859	8

association. Moreover, the MHC region has not emerged in analyses of other psychiatric disorders (eg, attention-deficit hyperactivity disorder, autism, bipolar disorder, major depressive disorder, and smoking behavior) using similar analytic methods and samples from overlapping sites. The MHC region emerges in only ~25% of diseases studied using GWAS (eg, multiple sclerosis (MS), rheumatoid arthritis, and systemic lupus erythematosus [SLE]). The association of genetic variation in the MHC with SCZ thus appears robust.

The Nature of the Beast

For the first time, we have a direct and empirical view of the genetic architecture of SCZ. The data tell a clear story—genetic variation for SCZ exists with frequencies from very common to rare. The risk conveyed by these variants is inversely associated with frequency. Empirical data about the allelic spectrum of risk for SCZ are depicted in Figure 1. In the upper left are rare CNVs of strong effect, common variants of quite subtle effect are on the lower right (red and yellow dots) and a polygenic signal (turquoise dots plus a blue best fit line).

The data that generated this graph provide an answer to a question that has bedeviled SCZ genetics for a century: is the syndromic entity SCZ a collection of rare Mendelian/Mendelian-like disorders or is it due to a large set of polygenes? SCZ is both.

The "map" of the genetic architecture of SCZ in Figure 1 is not the final draft. Larger GWAS will yield greater numbers of significant loci. More comprehensive analyses could yield more CNVs, and whole-genome, regional, and exome-sequencing efforts might add loci that are uncommon or rare.

What Lessons Have We Learned?

Very large samples are essential. Empirical data from >200 different human traits provide guidance for the determinants of the "success" of a GWAS. The single most important factor is sample size: very large samples by historical standards are required (a corollary is that negative results from small studies are meaningless). This relationship is illustrated in Figure 2 for SCZ.

Moreover, experience in human genetics indicates that if sample size is sufficiently large, many loci become

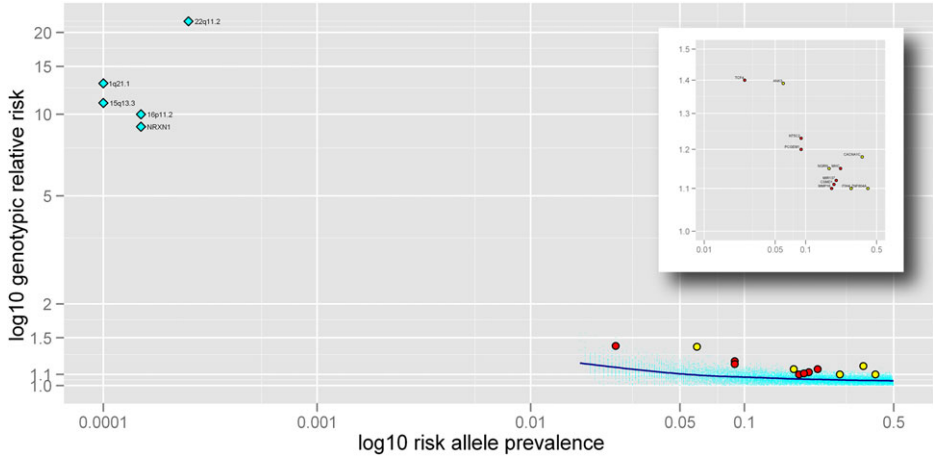

Fig. 1. The allelic spectrum of SCZ. Y-axis is genotypic relative risk, and x-axis is the risk allele prevalence (both $\log_{10}$ scale). In the upper left are rare CNVs of strong effect, common variants of quite subtle effect are on the lower right (red and yellow dots) and a polygenic signal (turquoise dots plus a blue best fit line). The inset expands the lower right to provide gene names.

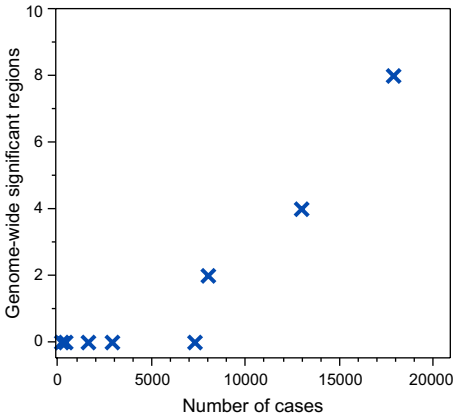

Fig. 2. Relation between number of genome-wide significant regions and number of cases for 8 GWAS of SCZ.

highly confident and pathways emerge that beautifully illuminate biology.[39,40]

Combining data across samples is valid. One argument that is not supported by data is that combining samples across different sites and countries will introduce crippling heterogeneity or bias. This argument has little support because there are dozens of examples where different studies have been combined to augment power. For example, height is surprisingly difficult to assess and yet a meta-analysis of 46 studies yielded compelling and coherent genetic results.[39]

Obey the laws of probability. Basic algebra classes include combinatorics and elementary probability. Application of these basic mathematical principles yields a conclusion of exceptional relevance to psychiatric genetics: gambling is not a strategy for progress. More specifically, everyone has to pay the price of multiple comparisons to avoid crippling type 1 error,[22] integrated replication is essential,[21] and underpowered studies are not worth doing. These principles are widely known but not universally appreciated.

Highly significant and replicated loci for SCZ typically have genotypic relative risks ~1.10. For replication of a specific association, the effect size in an initial study overestimates the actual value (ie, "winner's curse")[41] and high power is desired (90%). For one maker, 11 000 subjects are required (5500 cases and 5500 controls). Sample sizes increase to 17 500 subjects for 10 markers and 24 000 subjects for 100 markers. These sample sizes are more than an order of magnitude larger than historically typical for the SCZ candidate gene association field.

Hypothesis-driven approaches have not generally worked. In human complex disease genetics, empirical results have identified "usual suspects"[42] such as the MHC locus for type 1 diabetes mellitus (T1DM) or *APOE* for Alzheimer's disease. However, these examples are infrequent, and the vast majority of high-confidence results from genetic studies point to unsuspected loci. Therefore, SCZ researchers need to question seriously our most cherished ideas about the etiology of SCZ. We recently evaluated historical candidate genes for SCZ in comparison to GWAS findings and found essentially no overlap. The hypothesis-driven candidate genes for SCZ that have been studied the most had no indication of common-variant signal (ie, *COMT*, *DRD3*, *DRD2*, *HTR2A*, *NRG1*, *BDNF*, *DTNBP1*, and *SLC6A4*) (Collins et al, Submitted[10]). The status of "the special gene," *DISC1*, is particularly unclear: there is no common-variant signal, rare variation has not been found in resequencing of large samples, and reevaluation of the initial report suggests the name "disrupted in SCZ" is a misnomer. In this carefully assessed pedigree, the proposition had conduct disorder, and SCZ is a minor phenotype associated with the (1;11)(q42;q14.3) translocation (38% normal/other, 34% recurrent major depression, and 24% SCZ).[43]

What Should We Do Next?

In the complex and difficult history of SCZ genetics, many different technological approaches have been tried (table 1). To date, however, the only proven approaches are GWAS and assessment of CNVs. Should we push for more GWAS for SCZ?

First, as noted above, GWAS has an impressive track record of success in human complex disease genetics as well as in SCZ genetics. It has delivered where most other technologies have failed. Second, GWAS is a mature and inexpensive technology. Quality control, imputation, and analysis are readily accomplished. There are large amounts of data that can be used for comparisons. The cost can be as low as $250/sample. Third, most GWAS arrays contain single nucleotide polymorphisms (SNP) content sufficient to capture the majority of common variation in European and many other world populations along with content for large CNVs.

Fourth, we have empirical data that allow prediction of what might be discovered if we were to conduct GWAS on more SCZ cases. Given that research on the genetics of SCZ is 3–4 years behind other biomedical diseases, we can look at the data for SCZ in relation to other complex human traits instead of relying on assumption-laden predictions. We used the NHGRI GWAS catalog[23] to identify studies for 11 complex traits including SCZ. We reviewed 104 studies and included 72 (individual genome-wide genotyping of European subjects). Each study was reviewed at least twice to capture the number of cases analyzed and the number of genomic regions that exceeded genome-wide significance ($P < 5 \times 10^{-8}$). For each trait, we fit a regression line (number of regions ~ number of cases) to estimate the relation of these variables. There was a significant relation for 8 of the 11 traits.

The estimates in table 3 reveal intriguing lessons about the "mapability" of these 11 complex traits and provide

Table 3. Numbers of Cases and Significant Regions in 72 Studies of 11 Complex Traits

Trait	Heritability	Studies	Loci	Max N_{case}	Slope/1000 cases	*N*-to-first	*P* value
BMI	0.59	5	32	249 746	0.1	20 500	0.002
Breast CA	0.25	6	8	26 258	0.2	1050	0.05
T2DM	0.26	8	14	42 542	0.3	3273	0.0007
Lung CA	0.08	8	3	7560	0.4	3350	0.001
SCZ	0.81	9	8	17 836	0.4	4950	0.0004
AMD	0.46	4	7	6,777	0.5	50	ns
MS	0.41	8	6	4839	0.5	260	ns
Height	0.81	9	180	183 727	1.1	22 709	0.000005
T1DM	0.88	3	15	12 385	1.3	1592	ns
Crohn's	0.60	6	71	22 027	3.1	1248	0.01
SLE	0.44	6	12	2552	3.5	86	0.02

Note: Abbreviations. BMI, body mass index; CA, cancer; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; AMD, age-related macular degeneration; MS, multiple sclerosis; SLE, systemic lupus erythematosus. Slope is in units of number of number of genome-wide significant regions per 1000 cases. *N*-to-first is the number of cases required to get the first genome-wide significant finding. *P* value is the significance of the slope.

a glimpse into their genetic architectures. The slope estimates (ie, the number of genome-wide significant regions per 1000 cases) vary widely, from 0.1 for BMI to over 3 for Crohn's disease and SLE. Intriguingly, SCZ is in the middle of the pack and about the same as T2DM, lung cancer, age-related macular degeneration (AMD), and MS. However, when we estimate the number of cases required before the first genome-wide significant result, SCZ is the worst of the complex diseases in table 3 (the continuous traits of height and BMI are on a different scale). This is also apparent in Figure 2 where a "hockey-stick" relation can be imagined. The major correlates of the number of significant regions are sample size and heritability (Spearman ρ both ~0.6). Because heritability is not under experimental control, increasing sample size is therefore the way to increase the yield of GWAS (on average).

The estimates in table 3 encapsulate broad aspects of the genetic architectures of these traits (ie, the number of loci and effect size distributions as well as the inherent noisiness of phenotypic assessment). We do not see evidence that SCZ is qualitatively different save for a higher minimum number of cases to first detection (which we suspect reflects that there is no common genetic variant of strong effect as for AMD and T1DM combined with phenotype imprecision). SCZ is wonderfully typical.

What would happen if the numbers of SCZ cases were increased? In order to achieve the power of the most successful GWAS to data,[39] 50 000 SCZ cases and 50 000 controls are required (Ya[44] Around 12 000 SCZ cases with GWAS data are now available. If the sample size were increased to 100 000, there would be 2 immediate yields: the number of loci exceeding genome-wide significance can be predicted to total 21 and, equally important, the rank order of the SNPs would greatly improve meaning that pathway analyses would become far more reliable.

Continuing with GWAS would be one of the better bets for progress in the history of SCZ genetic research.

Should We Care if R^2 is Low?

For many complex traits, the amount of variance due to genome-wide significant loci (R^2) is a small portion of the overall heritability (<10%). Some have argued that this "missing heritability" is a reason why GWAS has failed and that their results do not matter.

We do not find this objection compelling. First, the goal of genetic studies of SCZ is to find pathways and loci that are strongly and robustly associated with disease. R^2 should not be the criterion for success because it is more relevant for individualized medicine (a distal not proximal goal). Second, we have remarkably poor heritability estimates for many complex traits meaning that this criterion is imprecise and subject to bias. Third, if sample sizes are too small, the genome-wide significance bar is conservative, and the impact of true effects that are not quite significant are missed. This issue is compounded by the fact that the current generation of GWAS SNP arrays imperfectly assess common genetic variation. If analyses account for these considerations, it can be seen that heritability is "hidden" rather than missing.[45,27,38] Indeed, 2 independent analyses suggest that common variants account for about a third of the variance in liability for SCZ.

Finally, some have argued that GWAS findings that do not have immediately obvious functional significance are irrelevant. This is a weak argument. Appropriate experimentation is required, and the literature is replete with examples of GWAS findings of mechanistic importance that emerged only after follow-up molecular work.

What About Sequencing?

There have been spectacular advances in sequencing technologies. It is now feasible to resequence all known exons

and even whole genomes. Costs are likely to decline but these are expensive technologies. At the time of this writing in Q1/2011, confident sequencing of an exome costs about $US 3000 and a genome costs $12 000. In the ideal situation, we would obtain genome resequencing for large collections of SCZ cases.

Does it make sense to shift direction entirely to sequencing as some have argued? It is unfortunate that sequencing has already been the focus of considerable "hype" (in contrast to GWAS where many investigators argued for conservative expectations).[46] There are reasons for caution, and we need to be mindful of painfully learned lessons from the past.

Psychiatric genetics has always underestimated the necessary sample sizes by several orders of magnitude. Although sequencing can discover causes of Mendelian disorders in small samples,[44] SCZ is unlikely to have Mendelian subforms. Quick successes are unlikely, and sequencing efforts could prove to be far more complex than predicted and very large samples are likely to be required. Exome and whole-genome sequencing works best if SCZ is caused exclusively or nearly so by rare variants that are not readily detected using the current generation of genotyping arrays. This strong assumption is inconsistent with empirical data.[38]

Sequencing is an appealing technology. However, we need to be realistic about what it might yield, and pay particular attention to its underlying assumptions and limitations. We need to be a bit jaded and worldly: many shiny new technologies have been applied to SCZ with great fanfare that ultimately failed to deliver. We should be interested, appropriately skeptical, and resist efforts to rely on a single technological approach.

The Take-Home

The main goal of genetic studies of SCZ is to identify pathways that confer risk and protection. For the first time, there is demonstrable progress toward this end as SCZ appears to be a relatively highly polygenic disease. If this knowledge is developed far more completely, we may be able to describe the basic mechanisms that go awry in the pathogenesis of SCZ (eg, the miR-137 hypothesis described above). Such knowledge can deliver compelling biological hypotheses to fuel more refined and specific investigations into the causes of SCZ.

Genetic approaches—particularly GWAS—are working. The field needs to stay focused on what has worked in order to maximize progress.

Funding

National Institutes of Health (NIH) grants (R01 MH080403, MH077139 to PFS); NIH Building Interdisciplinary Careers in Women's Health award (K12 HD01441 to SZ); NIH grant (T32 MH076694 to ST).

Acknowledgments

Conflicts of Interest: The authors report no conflicts. Author Contributions: All authors reviewed and approved the final version of the manuscript. PFS wrote the manuscript, YK coordinated the literature review, and YK, SZ, ST, and PFS conducted the literature review.

References

1. Psychiatric GWAS Consortium.Genome-wide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 2009;166:540–546.
2. Lichtenstein P, Bjork C, Hultman CM, Scolnick EM, Sklar P, Sullivan PF. Recurrence risks for schizophrenia in a Swedish national cohort. *Psychol Med*. 2006;36:1417–1426.
3. Lichtenstein P, Yip B, Bjork C, et al. Common genetic influences for schizophrenia and bipolar disorder: a population-based study of 2 million nuclear families. *Lancet*. 2009;373: 234–239.
4. Sullivan PF. The genetics of schizophrenia. *PLoS Med*. 2005;2:614–618.
5. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187–1192.
6. McGuffin P, Huckle P. Simulation of Mendelism revisited: the recessive gene for attending medical school. *Am J Hum Genet*. 1990;46:994–999.
7. Wray NR, Visscher PM. Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophr Bull*. 2010;36: 14–23.
8. Bassett AS, Chow EW, Weksberg R. Chromosomal abnormalities and schizophrenia. *Am J Med Genet*. 2000;97:45–51.
9. Ng MY, Levinson DF, Faraone SV, et al. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry*. 2009;14:774–785.
10. Collins AL, Kim Y, Sklar P, International Schizophrenia Consortium, O'Donovan M, Sullivan PF. Submitted. Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results.
11. Sanders A, Duan J, Levinson DF, et al. No significant association of fourteen candidate genes to schizophrenia in a large European-ancestry sample: implications for the field. *Am J Psychiatry*. 2008;165:497–506.
12. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Submitted. Genome-wide association study of schizophrenia identifies five novel loci.
13. International Schizophrenia Consortium.Greater burden of rare copy number variants in schizophrenia. *Nature*. 2008; 455:237–241.
14. Rujescu D, Ingason A, Cichon S, et al. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*. 2009;18:988–996.
15. Sebat J, Levy DL, McCarthy SE. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet*. 2009;25:528–535.
16. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008; 455:232–236.
17. Addington AM, Rapoport JL. The genetics of childhood-onset schizophrenia: when madness strikes the prepubescent. *Curr Psychiatry Rep*. 2009;11:156–161.

18. Levinson DF, Duan J, Oh S, et al. Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry*. 2011;168:302–316.
19. Murphy KC, Schizophrenia and velo-cardio-facial syndrome. *Lancet*. 2002;359:426–430.
20. Allen NC, Bagade S, McQueen MB, et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008;40:827–834.
21. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature*. 2007;447:655–660.
22. Sullivan PF. Spurious genetic associations. *Biol Psychiatry*. 2007;61:1121–1126.
23. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106:9362–9367.
24. Lander ES. Initial impact of the sequencing of the human genome. *Nature*. 2011;470:187–197.
25. Corvin A, Craddock N, Sullivan PF. Genome-wide association studies: a primer. *Psychol Med*. 2010;40:1063–1077.
26. Athanasiu L, Mattingsdal M, Kahler AK, et al. Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *J Psychiatr Res*. 2010;44:748–753.
27. International Schizophrenia Consortium.Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.
28. Lencz T, Morgan TV, Athanasiou M, et al. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry*. 2007;12:572–580.
29. Need AC, Ge D, Weale ME, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet*. 2009;5:e1000373.
30. O'Donovan M, Craddock N, Norton N, et al. Identification of novel schizophrenia loci by genome-wide association and follow-up. *Nat Genet*. 2008;40:1053–1055.
31. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009;460:753–757.
32. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744–747.
33. Sullivan PF, Lin D, Tzeng JY, et al. Genomewide association for schizophrenia in the CATIE study: results of Stage 1. *Mol Psychiatry*. 2008;13:570–584.
34. Ikeda M, Aleksic B, Kinoshita Y, et al. Genome-wide association study of schizophrenia in a Japanese population. *Biol Psychiatry*. 2011;69:472–478.
35. Kirov G, Zaharieva I, Georgieva L, et al. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry*. 2008;14:796–803.
36. Shifman S, Johannesson M, Bronstein M, et al. Genome-wide association identifies a common variant in the Reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet*. 2008;4:e28.
37. Sullivan PF. The Psychiatric GWAS Consortium: big science comes to psychiatry. *Neuron*. 2010;68:182–186.
38. Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol*. 2011;9:e1000579.
39. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010;467:832–838.
40. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010;42:937–948.
41. Ghosh A, Zou F, Wright FA. Estimating odds ratios in genome scans: an approximate conditional likelihood approach. *Am J Hum Genet*. 2008;82:1064–1074.
42. Burnett M, Alison J. Casablanca. Los Angeles, CA: Warner Brothers; 1942.
43. Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ. Schizophrenia and affective disorders–cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *Am J Hum Genet*. 2001;69:428–433.
44. Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. *Hum Mol Genet*. 2010;19:R119–R124.
45. Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genet Epidemiol*. 2010;34:254–257.
46. Psychiatric GWAS Consortium.A framework for interpreting genomewide association studies of psychiatric disorders. *Mol Psychiatry*. 2009;14:10–17.